

## Statistical Review and Evaluation

### Clinical Studies

NDA#: 20-498

Applicant: AstraZeneca

Name of Drug: Casodex (bicalutamide) 150 mg

Documents Reviewed: Volume 1 and Study Reports for Trials, 23, 24, and 25

Medical Officer: Scott Monroe, M.D. HFD-580

Statistical Reviewer: David Hoberman, Ph.D., HFD-715

### Background

The sponsor has submitted three (3) randomized, placebo-controlled, multi-center, parallel-group double-blind clinical trials in support of Casodex 150 mg as 'immediate hormonal therapy or as adjuvant therapy to treatment of curative intent in patients with non-metastatic prostate cancer.' Trial 0023 ( N=3292) was conducted in the US, trial 0024 (N=3603) in Europe, Mexico, and South Africa, and trial 0025 (N=1218) in Scandinavia.

Each trial began in 1995. The sponsor subsequently chose a data cutoff of June 2, 2000 allowing at least 2 years of follow up on each patient in all trials. The primary endpoint was time to objective progression as assessed by bone scan, X-ray, CT, MRI, ultrasound or biopsy. Other endpoints included time to treatment failure (essentially death, progression or withdrawal due to adverse event or switch to other cancer therapy), and the time to doubling of baseline PSA.

The Medical Division conducted several discussions with the sponsor concerning the possibility of bias in the detected time of progression arising from potential unblinding due to patient PSA levels and gynecomastia. *During these discussions, the Division suggested a binary endpoint of progression or death vs alive without progression, with progression confirmed by bone-scan.*

A potential problem with the clinical program which, in fact, becomes critical in the interpretation of the data, is that in the non-US trials (0024 & 0025), newly diagnosed patients are often put on "watchful waiting", i.e. not treated with cancer therapy, while in the US, virtually all patients undergo some treatment such as radical prostatectomy (RP) or radiotherapy. In addition, node positive patients were allowed in the non-US trials, but not in US trials. *Therefore, inferences about the efficacy of Casodex 150 mg in the non-US trials may not apply to the likely patient population in the US.*

### Sample Size and Protocol-Specified Analysis

Each trial's protocol-specified sample size was based upon slightly varying considerations. In 0023,

there was an anticipated 18% reduction in the hazard of progression (compared to placebo) based on a median time to progression of 7 years. The sponsor planned on a recruitment period of 2.5 years with a follow up of at least 2.5 years on each patient. Using essentially the same follow up scheme in 0024 and 0025, the anticipated reductions in the hazard were 20% and 30%, respectively. All resulting powers of tests using the planned Cox regressions were between 80% and 90%. However, the protocol for 0023 states the intention to combine the data from all 3 trials for the purpose of analyzing time to progression and survival. The covariates in the Cox model were to include baseline PSA, prior prostate cancer therapy, stage of disease, Gleason score, positive margins, and involvement of seminal vesicles.

## Trial Results

### **Baseline Characteristics**

**Tables 1 and 2** display the baseline characteristics of each treatment group in each of the 3 trials.

### **Primary Endpoint: Tumor Progression**

**Table 3** displays the results of each trial. Trial 0023 showed no evidence of efficacy, while the non-US trials achieved statistical significance below the .001 level for time to objective progression. *The endpoint analyzed by the sponsor includes **all detected progressions: bone-scan or 'other objectively confirmed', or death without progression.***

Note:

1) The overall incidence (regardless of sub-endpoint) in 0023 was considerably less than in the other trials,

2) There is no evidence at all that Casodex was associated with decreased mortality without progression. Moreover, **overall mortality** was 3.8% on Casodex and 3.7% on placebo in 0023 (US), 6.8% on Casodex and 7.6% on placebo in 0024 (Europe), and 11.4% on Casodex and 11.5% on placebo in 0025 (Europe). Figures are as of 2 June 2000. Thus there is no evidence that Casodex has an effect on death from prostate cancer.

3 *This reviewer has found that results in favor of Casodex in 0024 and 0025 are still significant if only **bone-scanned conformed progression** is analyzed using incidence analyzed by logistic regression or time to bone scan progression using a Cox model.* This result is important because it may eliminate any contention between the Agency and the sponsor concerning potential bias about timing of progression diagnosis and whether positive results rely on deaths and/or 'other objectively confirmed' progression. **Figure 1** displays the Kaplan-Meier plot of time to positive bone scan progression for 0023, indicating no evidence of treatment effect. **Figures 2 and 3** indicate longer times to bone scan progression for patients in the Casodex groups than the placebo groups in trials 0024 and 0025, respectively. Consequently, review of the progression data relies more on the issues of 1) the applicability of positive findings in non-US trials to the US patient population, and 2) validity of the bone-scan reading addressed in separate re-reading study conducted by the sponsor.

### **Evidence of Efficacy in Subgroups (0024 and 0025)**

Given the striking difference in results, FDA was particularly concerned to know whether the foreign

studies provided evidence of efficacy in subgroups. For example, the table below displays the percentage of patients who received RP, Radiotherapy or Watchful Waiting in trials 0024 and 0025.

Radical Prostatectomy Radiotherapy Watchful Waiting

0024 46% 18% 36%

0025 13% 6% 81%

The table below displays the percentage of patients in each treatment group who had **bone-scan confirmed progression (BSCP)** in each prior therapy category above:

0024

Casodex 2% 5% 4%

Placebo 4% 10% 7%

0025

Casodex 1% 10% 5%

Placebo 13% 35% 15%

The results suggest that in the foreign patients, some benefit may have occurred regardless of prior therapy (or lack thereof). It appears that, overall, patients who got radiotherapy, only, did not do as well as those on watchful waiting. However, it should be noted that the percentages reported for radiotherapy are based on denominators in the range of 30, while those for watchful waiting have denominators in the range of 500. More likely is the possibility that patients with watchful waiting have less severe disease so that the incidences are not comparable.

**Lymph node status was also examined but no useful inference can be drawn since virtually all patients in both non-US trials either did not have positive lymph nodes or regional lymph nodes could not be assessed.**

**Table 4** displays the sponsor's subgroup analyses for 0024 and 0025 using all 3 sub-endpoints for progression. Generally, trends in favor of Casodex over placebo exist in the prognostic subgroups.

**Comparability of non-US to US Patients Samples**

The sponsor accounts for the null result of trial 0023 by stating that "in these patients, watchful

waiting was not an option for prior therapy, the patients were of younger age, had low PSA values, and were not eligible for entry into the trial if they had known positive nodes. *Pre-procedure* [emphasis added] PSA levels were lower in trial IL0023 than in the other 2 trials suggesting that a lower tumor burden was evident in these patients from the outset. Collectively, these facts explain the relative immaturity of Trial IL0023 in terms of TTP [time to progression], with just 5.0% and 5.3% of CASODEX- and placebo-treated patients, respectively, having progressed or died at the time of data cut-off." (ISE, p.60)

This reviewer has conducted a logistic model analysis which assesses the relation between baseline prognostic covariates and incidence of BSCP in pooled trials 0024-0025 (combining both treatment groups). Gleason Grade was not used because the Medical Division concluded that the grades were assigned differently in Europe than in the US. For instance, **Table 2** indicates that Gleason scores tended to be lower in the US than Europe, yet the incidence of progression in the US was much lower. Tumor class was stratified by T1/T2 vs T3/T4. Previous therapy was stratified by having had *at least* RP vs only radiation or neither therapy. Age was not a contributing factor in the model.

Assuming independent contributions of each risk factor (i.e., without interactions with each other or treatment), it was found *that the odds of bone-scan progression were multiplied by the following factors* relative to the following defined index subgroups:

**Tumor Class:** being in T3/T4 rather than T1/T2: odds multiplier=3.4.

**Previous Therapy:** only radiation or neither therapy vs at least RP: odds multiplier=2.6. **PSA:** When restricting patients to those who had at least RP (the cohort of interest in comparing results from US and Europe), pre-procedure PSA was found to be a significant prognostic variable for BSCP.

Since:

1. The percentage of patients in T3/T4 who had at least RP in the US was 27% while the respective figure in Europe was 43%,
2. The percentage of patients with at least RP in the US was 80% and in Europe was 38%, and
- 3) The distribution of pre-procedure PSA's was somewhat less in the US than in Europe,

then it is not surprising that the **overall** incidence of BSCP in the US was less than in Europe.

Having made this simple observation, it is natural to ask: "How much of the discrepancy in BSCP incidence can be explained by known or potential prognostic factors whose distributions differ between the two locations? In particular, what incidences in the two treatment groups would one expect in the US using the data supplied from Europe?" Of course, a projection of this kind must assume that the relationships between the prognostic variables and BSCP are similar in the US and Europe, and that baseline measurements were conducted similarly. With regard to the former, a logistic model yields similar coefficients for the contribution of tumor class and pre-procedure PSA in the two regions. However, given the Medical Division's judgement about the unreliability of the Gleason Grades, the latter assumption is problematic.

Nevertheless, this reviewer performed an exploratory analysis as follows:

1. Restrict the analysis to (European) patients in T3/T4 **and** who had at least RP as previous

- therapy (N=802, # of events = 49 in patients with pre-procedure PSA's).
2. Find the quartiles of pre-procedure PSA in these patients and then determine the incidence of BSCP in each treatment group for each of the quartiles.
  3. Compute the percentage of patients in the US (cohort of T3/T4 and at least RP, N=850) who fall into each of the European-determined quartiles.
  4. Finally, use the law of total probability to compute the *expected* incidence of BSCP in both treatment groups in the US.

Restricting the analysis to T3/T4 is a way to control for the unequal distribution of that tumor class stratum between Europe and the US. Although this restriction obviously does not account for all the events, it does address a population which may get substantial benefit from the drug. *Specifically, the European study is handicapped by the fact that out of the total of 303 BSCP's, only 66 occurred in patients with at least RP, thus limiting its utility in projecting efficacy to the US.* Nevertheless, restricting patients to those with at least RP almost completely addresses the issue of the relevant population in the US.

The quartiles of pre-procedure PSA in the European subgroup of interest were: 25% ile=8.1  $\mu$  g/L, median=12.5  $\mu$  g/L, and 75% ile=22  $\mu$  g/L

The table below displays the incidences in each pre-procedure PSA quartile stratum for steps 2 and 3, above

1<sup>st</sup> 2<sup>nd</sup> 3<sup>rd</sup> 4<sup>th</sup>

Incidence of BSCP in 0024-25: Casodex 1.7% 1.1% 5.5% 6.6%

Incidence of BSCP in 0024-25: placebo 6.7% 12% 5.9% 10%

Quartile percentage in 0023 51% 25% 16% 8%

The projected incidences are 8.1% in the US placebo group and 2.3% in the Casodex group. Thus,

there is virtually no reduction in incidence with adjustment for pre-procedure PSA from the actual European incidences of 8.7% and 3.6%, respectively. In the US, both treatment groups had an incidence of only 1.5%-1.8%. When "**other objectively confirmed**" progressions are included, the results are similar, with the projection in the US being 15% in the placebo group and 4% in the Casodex group. The *actual* US incidences were 3.7% in the placebo group and 3.1% in the Casodex group. Thus, pre-procedure PSA may be a risk factor, but the disparity between the distributions in Europe and the US is not large enough to induce any substantial difference in risk within the highest tumor classes. There is also a curious phenomenon in the above table. Note that the treatment difference emerges not from the contribution in the *higher* PSA quartiles, but in the *lower* ones, counter to what one might expect. Of course, working with small numbers of events may create some instability in estimates. Another interesting fact is that only 40% of the BSCP's in the US occurred in T3/T4 (in the at least RP cohort), where 70% in Europe occurred in T3/T4. It is possible that a significant number of US patients were assigned to T3/T4 when they should have been assigned to T1/T2.

Finally, we may review the explanation for the 'low' incidence of BSCP in the US cited from the sponsor's ISE at the beginning of this section. From

1. The above exploratory analysis using **tumor class**,
2. The fact that virtually all BSCP's occurred in **node negative patients** in the US and Europe,
3. The weak effect of **pre-procedure PSA** to explain differential risk of objective progression, and
4. The fact that **age** is not a prognostic factor for BSCP,

we can fairly say that *although the sponsor's hypothesis is a plausible explanation for the US trials negative results, the data is not consistent with that hypothesis*. This does **not** mean that the thrust of the sponsor's argument is wrong. It is just that one cannot use the data from the trials to support it in any convincing way. In fact, the results from the projection raise concerns that the baseline data are not useful for any analytical purpose due to the possibility that assignments of tumor classes and Gleason grades were not consistent between the US and Europe.

Note that the foregoing discussion considered *pre*-procedure PSA because that is one factor that the sponsor held responsible for the difference in incidence between Europe and the US (see quote from ISE, above). However, pre-randomization (post-procedure PSA) is likely to be more prognostic than pre-procedure PSA for the obvious reason that the goal of RP is to lower PSA as much as possible. If there is still residual tumor, then post-procedure PSA's may still be elevated and the risk of progression greater in those patients.

With that in mind, we note that, in the US, the minimum value for post-procedure PSA is 0.2  $\mu$  g/L , whereas the minimum in Europe is 1.0  $\mu$  g/L. If these are taken to be the limits of detection, we can stratify all patients into two strata: above 1.0  $\mu$  g/L or at most 1.0  $\mu$  g/L. The table below displays, as did the previous table, the incidences of BSCP in each of the 2 strata in Europe and the percentage of patients falling into those strata in the US:

Post-procedure PSA

<=1.0 >1.0

Incidence of BSCP in 0024-25: Casodex 1.5% 15.2%

Incidence of BSCP in 0024-25: placebo 5.7% 19.2%

Quartile percentage in 0023 95% 5%

The projected US incidence in the Casodex is now 2.2% and in the placebo group, 6.4%, not much of an improvement over using the pre-procedure PSA.

#### Sponsor's Subsequent Sub-setting

In response to questions submitted to the sponsor during the review period, the FDA received a submission on 5/13 narrowing the patient population which may derive benefit from Casodex 150 mg in the US. One sub-population is patients who undergo radical prostatectomy with or without radiotherapy. The sponsor used modeling techniques to narrow this group to patients who had T3/T4 disease together with 'detectable' post-procedure PSA, in this case  $> 0.2 \mu\text{g/L}$ . The sponsor's results are displayed in the table on the next page.

#### **Comparative findings for Casodex vs placebo for rate of progression events among subpopulations of patients previously treated with radical prostatectomy**

Subpopulation characteristics	N	Casodex	Placebo	Hazard ratio
(Radical prostatectomy patients)		% patients who progressed		(95%CI)
<b>Postsurgical PSA</b>				
Nondetectable ( $0.2 \text{ ng/ml}$ )	3380	3.5	4.9	0.69 (0.50, 0.97)
Detectable ( $>0.2 \text{ ng/ml}$ )	932	11.2	17.4	0.61 (0.43, 0.86)
<b>Detectable postsurgical PSA, by disease stage</b>				
Localized ( $T_{1-2}$ ), PSA $>0.2 \text{ mg/ml}$	423	9.2	11.2	0.74 (0.40, 1.37)
Locally advanced ( $T_{3-4}$ ),	509	12.9	22.3	0.54 (0.35, 0.83)
<b>PSA <math>&gt;0.2 \text{ ng/ml}</math> Detectable postsurgical PSA and locally advanced disease, by trial</b>				

Trial 23	158 (6%) <sup>a</sup>	9.6	16.0	0.53 (0.21, 1.37)
Trial 24	277 (16.8%) <sup>a</sup>	15.0	24.3	0.55 (0.32, 0.96)
Trial 25	74 (46.5%) <sup>a</sup>	12.1	26.8	0.49 (0.15, 1.58) <sup>b</sup>

<sup>a</sup> Percentage of radical prostatectomy patients with locally advanced disease and detectable postsurgical PSA per trial.

<sup>b</sup> Fewer than 20 events.

-

By reporting the results of the individual studies in the lower part of the table, the sponsor is trying to convey some version ‘consistency’ of the hazard ratios for progression (the sponsor has included BSCP, other objective progression, and death) from trial to trial. In other words, if one looks at the subset derived from their data searching, then the hazard ratio for progression in the US (Trial 23) and Europe are similar (about 0.50), albeit with a wide confidence interval. Of course, this is not a ‘statistical’ validation of similarity, but a pattern than falls out of the data.

One problem with this strategy is that the comparison of 9.6% vs 16% in Trial 23 is based upon only 20 events. Another is the fact that *if only BSCP progressions are counted, there is no trend in favor of Casodex: 5 in the Casodex group and 4 in the placebo group.*

-

-

-

-

-

-

**The Centralized Scan Re-Read Study**

In all trials, local investigators evaluated bone scans as either positive or negative for metastasis. At the request of the FDA, the sponsor conducted a study in which *all positive scans and a sample of negative scans* were re-read by an independent committee of evaluators. Central readers had access only to the baseline scan and follow-up scan in order to conduct a ‘paired read’. No clinical data was used.

*The primary objective of the study was to re-evaluate the extent to which locally read negative scans were read as positive by the central re-read (negative to positive conversion).* Missing and indeterminate scans were either allocated as positive in a proportionate manner, i.e. according to the

overall positive proportion, or as in an alternative scheme, all indeterminate scans were counted as negative, and missing scans were distributed proportionately. Using that information, a projected estimate of the total number of positive scans was computed for each treatment group. Confidence intervals for the estimated number of positive scans was then computed, and finally, a range of estimated odds ratios combining the 3 trials were computed comparing Casodex to placebo with respect to the projected probability of a positive scan. *The sponsor never intended to conduct a statistical test for treatment effect using the projected number of positive scans.*

**Table 5** displays the sample sizes for each study required to provide confidence intervals for the probability of conversion from negative to positive with the specified precision. For purposes of sample size calculation, the sponsor assumed a conversion probability of 5%. For example, referring to the last column, the chosen sample size for the study provides a 2-sided 95% confidence interval whose precision is +/- 3%.

The final sample size was 1459 patients, of which 1336 (91.6%) had had follow-up scans performed. Ultimately, 89.8% of the patients sampled for the study had both a baseline scan and follow-up scan available for the central paired read. Thus some follow-up scans were read without a baseline scan.

**Table 6** displays cross-tabulation results correlating the local reads to the subsequent central reads combining all trials, while **Table 7** stratifies **Table 6** by trial. Of the 1459 sampled patients (patients with or without scans), 1130 had follow-up scans that could be evaluated as either negative or positive with the remaining 329 scans either "indeterminate" or missing. After review of the design, analytic plan and results of the study, it became clear to this reviewer that as long as indeterminate and missing scans were not allocated as positive differentially between the treatment groups, their role is minor.

**Table 8** displays the main results of the study including the negative to positive conversion proportions in each trial. **This table excludes patients with missing or indeterminate scans.** Note that the local read negative to central read positive conversion proportions are all between 4% and 7%, very close to the sponsor's assumption of 5%. *In addition, there is no clear evidence that these proportions differ between the treatment groups.*

### Final Results

The major result of the entire study is displayed in **Table 9**. The far left-hand column designates analyses which included various proportions of *non-bone scan progression* included in the projections. However the result most germane to the FDA is the last row which includes only bone scans. Thus, the result is an odds ratio of a little over .80 for the central read, while the far right column indicates that the result of the local read was an odds ratio of .73. **Table 10** displays the bone scan, only, results when the upper and lower bounds of the 90% confidence interval for the total number of positive scans is used. The sponsor simply asserts that point estimates of the projected odds ratio still favors Casodex for objective progression and deliberately avoids the role of standard error of any estimate.

### Reviewer's Analysis and Discussion

The purpose of this section is to explain the circumstances under which the results of this study *could have cast* substantial doubt on the statistical results of the treatment comparisons using the original local reads; for it is not clear what the re-read study was supposed to accomplish. There could be at least two roles of a truly non-zero negative to positive conversion proportion: It can be used to relate

the expected treatment difference using the central read to the original local read treatment difference (see below), and 2) it could be used by itself as a measure of uncertainty about the consistency of local-central reads. This section concentrates on 1) because the relative efficacy of the two treatments is the paramount question.

Begin with the following considerations:

Let  $\Pr_{++} = \Pr(C_+|L_+)$ , the probability that the central read of a random scan is read as positive given that the local read is read as positive. Also, let  $\Pr_{-+} = \Pr(C_+|L_-)$ . Using the law of total probability for a positive scan:

1.  $\Pr_D(C_+) = \Pr_{++} \Pr_D(L_+) + \Pr_{-+} \Pr_D(L_-)$  is the probability of a positive scan in the drug group. Again,  $\Pr_{-+}$  is the negative to positive conversion proportion from the local to the central read. A similar expression for the placebo group gives:
2.  $\Pr_P(C_+) = \Pr_{++} \Pr_P(L_+) + \Pr_{-+} \Pr_P(L_-)$

The reason that the conditional probabilities do not have treatment identifiers is that **there is no reason to believe that these probabilities are different for the Casodex and placebo groups as long as the reads are blinded to treatment assignment**. In fact, the data give no indication of bias. Importantly, the issue of differential conversion probabilities was not addressed in the design of the study. Such an "interaction" with treatment would have provided a substantial source of bias, but, as we see, there is no reason for believing that such a bias exists. After algebraic simplification, the differences between treatments between the two reads is expressed by:

$$3. \Pr_P(C_+) - \Pr_D(C_+) = [\Pr_{++} - \Pr_{-+}] \times [\Pr_P(L_+) - \Pr_D(L_+)]$$

Equation 3) shows that if  $\Pr_{++} = \Pr_{-+}$ , then the treatment difference in the central read disappears because of multiplication by zero on the right side. This is the case in which the readings by the central readers are statistically independent of those of the local read. Thus, we would hope that  $\Pr_{++} \gg \Pr_{-+}$ . And in fact, this is so in this data:  $\Pr_{++}$  is in the range of 70%-75% in 0024 and in the 90%'s in 0025.  $\Pr_{-+}$  is in the range of 7% in all studies. Thus, there will *always* be *some* decrease in the central read treatment difference compared to the local read difference as long as  $\Pr_{-+} > 0$  or  $\Pr_{++} < 1$ . In terms of statistical testing, this shrinkage toward the null is not important as long as  $\Pr_{++} - \Pr_{-+}$  is known, because that factor cancels out when a z-statistic is formed. The analogue to logistic regression is that the odds ratio *must* shrink toward 1, the null value, but as in the case of a difference between proportions, the standard error decreases in tandem.

As an example, this reviewer has done the calculations for trial 0024 using **Table 7** and **Table 8**. The original local read percentage of positive bone scans in the placebo group was  $60/1798 = 3.3\%$  with  $116/1805 = 6.4\%$  in the Casodex group. The difference is 3.1%. However the projected central read percentage in the placebo group is 12.5% while that in the Casodex group is 10.0%, a difference of 2.5%. In other words, it was inevitable that the treatment difference would decrease as long as any previously negative scans were converted to positive scans by the central readers. *Similarly, it was inevitable that the sponsor's original odds ratio would shift towards the null.*

These considerations lead this reviewer to conclude that, from the beginning, there was little likelihood that the re-read study would add anything substantial to the data which already existed. It does, however, provide information that may reflect upon whether or not there was adequate blinding of the local readers. If local readers had read placebo scans positive in a biased manner, then we would expect the central readers to have a "high" positive to negative conversion proportion. But this is not the case: those proportions in the Casodex and placebo groups are 23.2% and 20.0%, respectively, thus providing no evidence that the local readers were biased in reading placebo scans. It provides no substantial evidence that the sponsor's analyses of the local reads are in doubt as far as the validity of the original analyses submitted to the FDA. Since, apparently, no one envisioned the re-read study as a source of another statistical test of treatment effect, then the question of the variance of the left side of equation 3) when  $Pr_{++} - Pr_{-+}$  is a random variable is not an issue. In any case, this reviewer has determined that the additional variance introduced by estimating  $Pr_{++} - Pr_{-+}$  in this study is negligible.

### **Conclusions**

-

Each European trial provides statistically significant evidence that Casodex 150 mg delays or possibly prevents objective progression of early prostate cancer as measured by bone scans positive for metastases. However, the inclusion of disparate patient groups which received different background therapy (or lack thereof), complicates inference by inviting examination of treatment comparisons within subgroups of patients. Nevertheless, there is evidence in each European trial that patients who either underwent previous therapy (radiation and/or radical prostatectomy) or who underwent 'watchful waiting' derived some benefit from Casodex therapy. This evidence takes the form of 1) consistent direction of effect for Casodex in clinically relevant subgroups in both trials and of (although not rigorous statistical evidence) and 2) nominally low p-values *in both trials separately* comparing Casodex to placebo in each of these clinically relevant subgroups. However, based on Casodex labeling in other countries and the recent data- searching by the sponsor, the revised version of the sponsor's indication now excludes patients who underwent RP with T1/T2 disease. (See table on page 8).

However, there is no evidence that Casodex would be beneficial to patients who underwent previous therapy in the US (the only class of patients who were studied in the US). Moreover, the sponsor has not provided evidence that 'watchful waiting' patients would benefit from Casodex 150 mg in the US. In fact, there is reason to believe that patients who got 'watchful waiting' in Europe are not the same patients who would receive 'watchful waiting' in the US. Although the sponsor claims that the patient population undergoing previous therapy in Europe and the US were quite different, adjusting for clinically relevant factors at baseline does not seem to account for the difference in incidence, a sign that baseline measurements may not be calibrated between the two regions, associations with progression are weak, and/or that there are some unrecorded factors which depressed "objectively confirmed progression" in the US. It is also possible that differences in clinical practice rendered the US study such that the prostatectomy patients were fated to get no benefit from Casodex. Only monitoring the current US trial can answer that question. Given the lack of sufficient information in the US at this time, one course is to request further data from the Trial 23 so that a decision does not rest solely on an extrapolation of results from Europe.

David Hoberman, Ph.D.

Mathematical Statistician

Concur: Dr. Welch

Dr. Nevius

cc:

Arch NDA# 20-498

HFD-580

HFD-580/SMonroe, DShamus, DSpell-LeSane

HFD-715/DHoberman, MWelch, ENevius, CAnello

## Tables and Figures

**TABLE 1**

### **BASELINE DEMOGRAPHIC CHARACTERISTICS**

Demographic characteristic	IL0023		IL0024				IL0025	
	CASODEX (N=1647)	Placebo (N=1645)	CASODEX (N=1798)	Placebo (N=1805)	CASODEX (N=607)			
Age (years)								
Mean	64.5	64.4	68.6	68.7	68.5			
SD	7.02	7.08	7.29	7.13	5.16			
Range	42 to 85	38 to 83	42 to 93	46 to 93	46 to 87			
Age distribution (n [%])								
<55 years	151 (9.2)	155 (9.4)	62 (3.4)	51 (2.8)	10 (1.6)	4		
55 to <65 years	614 (37.3)	607 (36.9)	422 (23.5)	432 (23.9)	99 (16.3)	113		
65 to <75 years	780 (47.4)	785 (47.7)	936 (52.1)	934 (51.7)	475 (78.3)	468		
≥75 years	102 (6.2)	98 (6.0)	378 (21.0)	388 (21.5)	23 (3.8)	26		
Weight (kg)	n=1536	n=1526	n=1742	n=1738	n=602			

Mean	85.38		84.60		77.28	78.08		79.31			
SD	14.21		13.90		11.35	11.75		11.97			
Range	49 to 166		46 to 160		45 to 132	40 to 135		46 to 143			
Race (n [%])											
White	1369	(83.1)	1391	(84.6)	1714	(95.3)	1709	(94.7)	606	(99.8)	607
Black <sup>a</sup>	191	(11.6)	188	(11.4)	17	(0.9)	13	(0.7)	0	(0.2)	0
Other <sup>b</sup>	87	(5.3)	66	(4.0)	67	(3.7)	83	(4.6)	1	(0.2)	4

<sup>a</sup>Includes Afro-Caribbean.

<sup>b</sup>Includes Asian, Oriental, Hispanic, Mixed race.

N number of patients randomised.

n number of observations (if less than N).

## TABLE 2

### Baseline disease characteristics: individual trial data

Characteristic	IL0023		IL0024		PI
	CASODEX 150 mg (N=1647)	Placebo (N=1645)	CASODEX 150 mg (N=1798)		
Tumour stage: T category (%) <sup>a</sup>					
T1	158	(9.6)	160	(9.7)	45
T2	1033	(62.7)	1040	(63.2)	74
T3	452	(27.4)	442	(26.9)	56
T4	4	(0.2)	3	(0.2)	46
TX	0		0		0
Gleason score (%)					
Well differentiated (2,3,4)	69	(4.2)	79	(4.8)	56
Moderately differentiated (5,6)	789	(47.9)	798	(48.5)	74
Poorly differentiated (7,8,9,10)	789	(47.9)	768	(46.7)	47
Not known	0		0		28
Lymph node category (%)					
N-	1186	(72.0)	1172	(71.2)	10
N+	1	(0.1)	0		48
NX	460	(27.9)	473	(28.8)	66

Previous therapy of curative intent (%) <sup>b</sup>							
Radical prostatectomy <sup>c</sup>	1322	(80.3)	1325	(80.5)	835	(46.4)	81
Radiotherapy only <sup>d</sup>	325	(19.7)	320	(19.5)	335	(18.6)	32
Watchful waiting	0		0		628	(34.9)	66
Other <sup>e</sup>	0		0		0		1

<sup>a</sup>Represents a mixture of clinically or pathologically staged specimens.

<sup>b</sup>Mutually exclusive categories.

<sup>c</sup>Includes radical prostatectomy with radiotherapy.

<sup>d</sup>Includes brachytherapy. <sup>e</sup>Includes cryotherapy/cryosurgery.

TX/NX tumour stage/lymph nodes not assessable. N- No regional lymph node metastasis. N+ Includes categories N1, N2, and N3 (metastasis in lymph node [local or regional]). N number of patients randomised.

TABLE 3

### Summary of patients with disease progression

Type of Progression	Number (%) of patients											
	IL0023		IL0024				IL0025				Combined data	
	CASODEX (N = 1647)	Placebo (N = 1645)	CASODEX (N = 1798)	Placebo (N = 1805)	CASODEX (N = 607)	Placebo (N = 611)	CASODEX (N = 4052)	Placebo (N = 4061)				
Objective <sup>a</sup>												
Death <sup>b,c</sup>	52 (3.2)	55 (3.3)	96 (5.3)	92 (5.1)	48 (7.9)	44 (7.2)	196 (4.8)	191 (4.7)				
Bone scan <sup>b,d</sup>	21 (1.3)	15 (0.9)	60 (3.3)	116 (6.4)	32 (5.3)	95 (15.5)	113 (2.8)	226 (5.6)				
Other <sup>b,e</sup>	10 (0.6)	17 (1.0)	25 (1.4)	85 (4.7)	19 (3.1)	40 (6.5)	54 (1.3)	142 (3.5)				
<b>Total</b>	<b>83 (5.0)</b>	<b>87 (5.3)</b>	<b>181 (10.1)</b>	<b>293 (16.2)</b>	<b>99 (16.3)</b>	<b>179 (29.3)</b>	<b>363 (9.0)</b>	<b>559 (13.8)</b>				
Non-objective <sup>f</sup>	0	0	5 (0.3)	31 (1.7)	12 (2.0)	53 (8.7)	17 (0.4)	84 (2.1)				

<sup>a</sup> Includes death in the absence of objective progression (see Section 3.2.1).

<sup>b</sup> Categories are mutually exclusive.

<sup>c</sup> In the absence of objective progression.

<sup>d</sup> Bone-scan-confirmed progression.

<sup>e</sup> Other objectively confirmed progression, eg, magnetic resonance imaging, computerised tomography, biopsy.

<sup>f</sup> Patients with positive subjective assessments but no positive objective confirmation of progression. Patients with subjective

TABLE 4

**Summary of TTP events for immediate therapy patients (Trials IL0024 and IL0025<sup>a</sup>)**

Subgroup	Events (%) in CASODEX arm <sup>b</sup>		Events (%) in placebo arm <sup>b</sup>		Hazard ratio <sup>c</sup>
	Events	(%)	Events	(%)	
All watchful waiting patients	172/1114	(15.4)	286/1171	(24.4)	0.53
by tumour stage <sup>d</sup>					
localised	103/779	(13.2)	154/848	(18.2)	0.65
locally advanced	69/335	(20.6)	132/323	(40.9)	0.42
by Gleason score					
well	62/501	(12.4)	82/521	(15.7)	0.64
moderate	59/424	(13.9)	124/454	(27.3)	0.43
poor	48/164	(29.3)	75/176	(42.6)	0.58
by PSA category					
0 to 4µg/l	17/169	(10.1)	19/193	(9.8)	1.08
>4 to 10µg/l	33/267	(12.4)	39/240	(16.3)	0.65
>10 to 20µg/l	48/275	(17.5)	59/278	(21.2)	0.68
>20µg/l	69/376	(18.4)	160/428	(37.4)	0.38

<sup>a</sup> Watchful waiting patients were ineligible for Trial IL0023

<sup>b</sup> Events are objectively-confirmed progression or death in the absence of progression in all 3 trials.

<sup>c</sup> Hazard ratio is for CASODEX versus placebo.

<sup>d</sup> Patients with locally advanced disease are categorised as T3, T4, TX or N+; all other patients are considered to have localised disease

TABLE 5

Trial Number not Sample size per randomized Probability of Probability of

Objectively treatment group to determine determining the determining the

Progressed reclassification rate with a reclassification rate reclassification rate

SE of approx. 1.5% to within  $\pm$  2.5% to within  $\pm$  3%

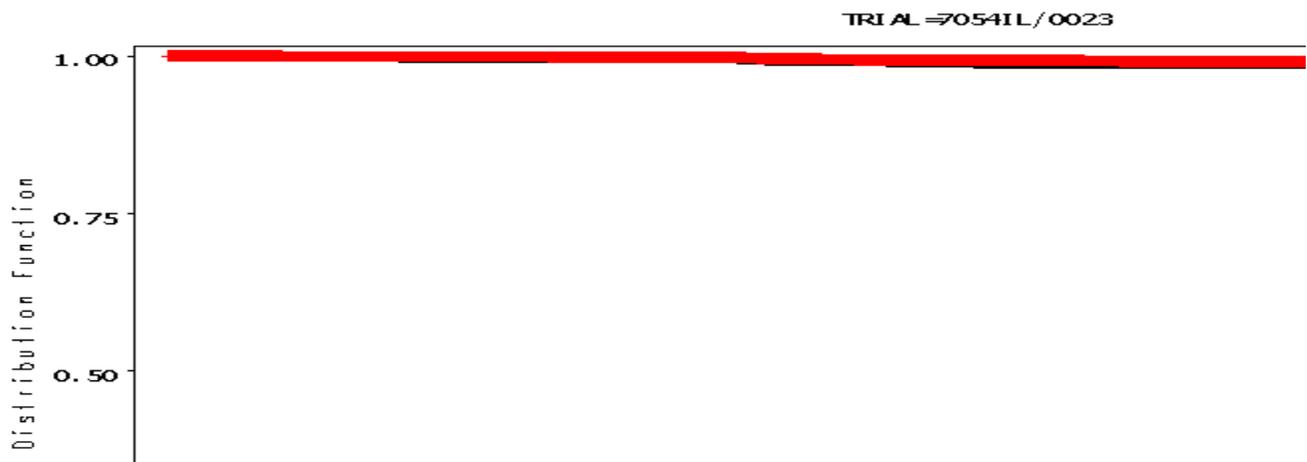
0023 3122 180 90% 95%

0024 2129 180 90% 95%

0025 940 145 90% 95%

Allowing 10% overage for those patients in whom no 2 year bone scan was taken, a total of 400 patients will therefore be sampled from trial 0023, 400 from trial 0024 and 320 from trial 0025.

FIGURE 1



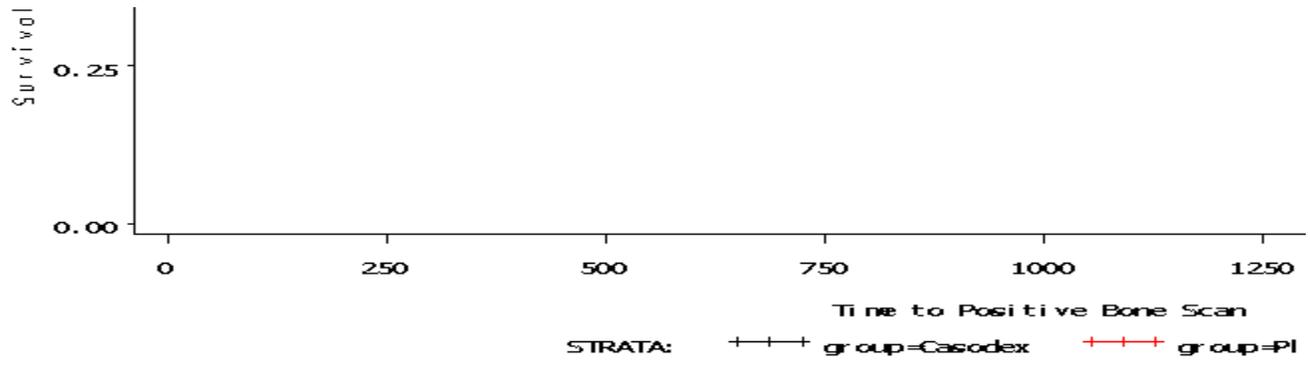


FIGURE 2

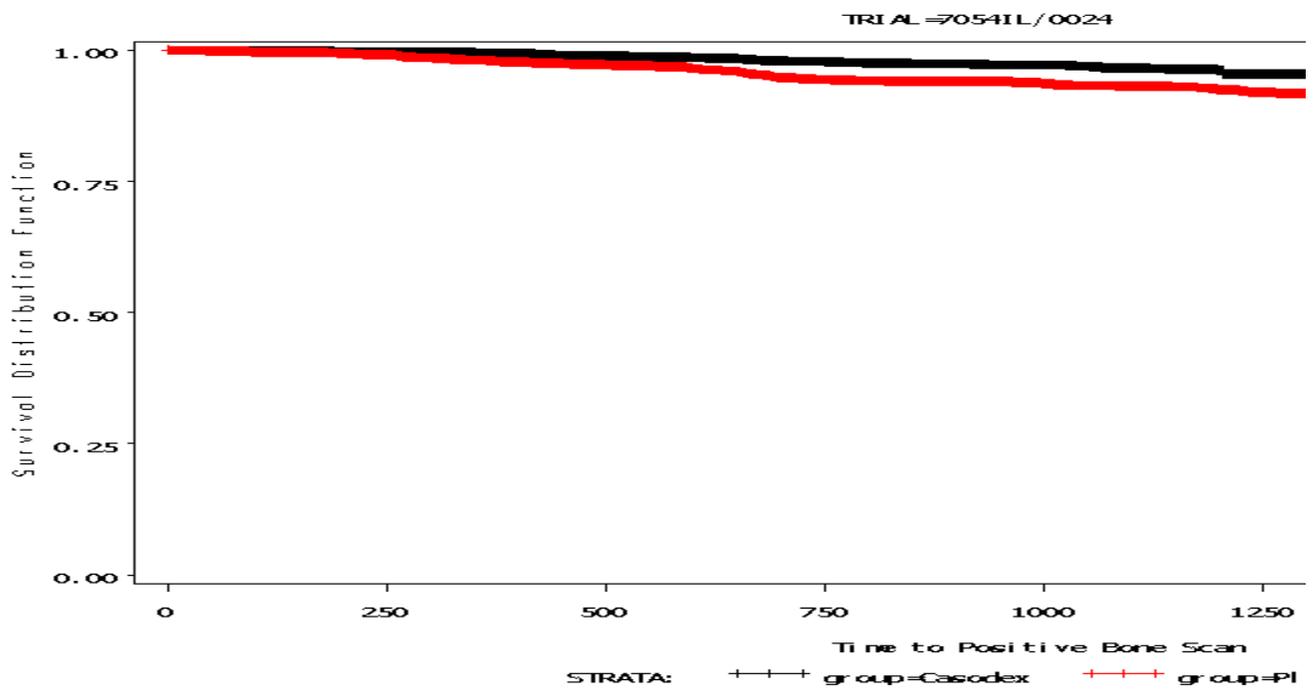
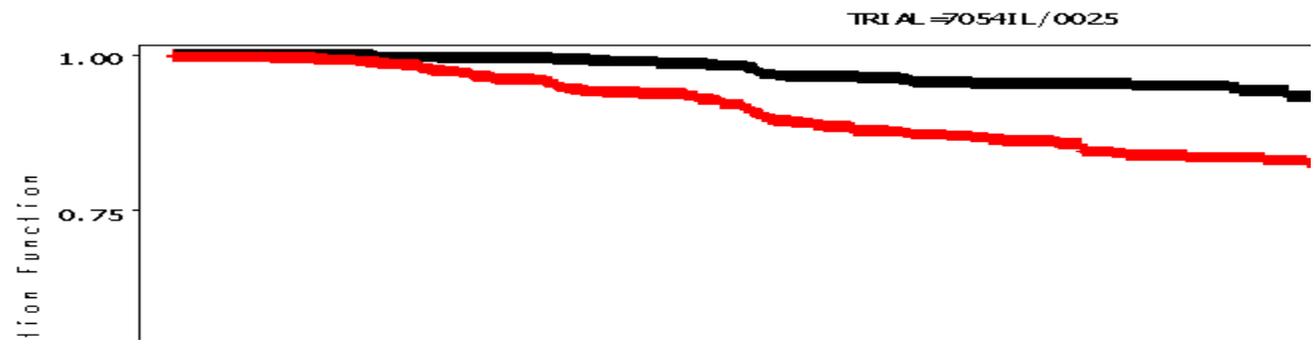
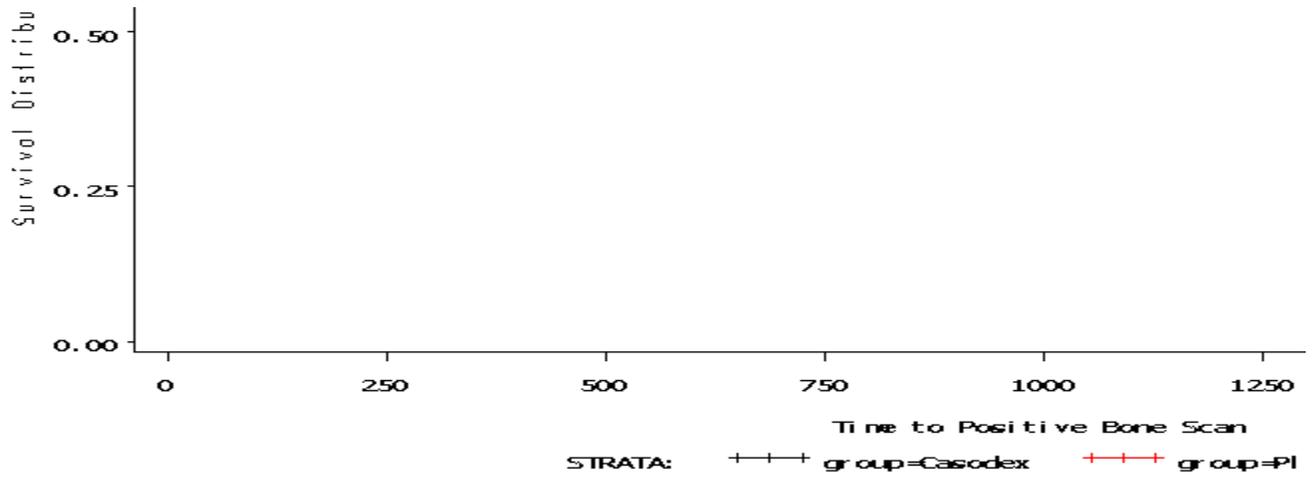


FIGURE 3





**TABLE 6 Distribution of outcomes for the central re-read of the follow-up scans, overall and by randomised treatment**

Treatment group	Local follow-up result	Number of patients	Number (%) of patients with central follow-up result		
			Outcome of + or -	Indeterminate	Missing
CASODEX	+	113	95 (84.1)	5 (4.4)	13 (11.5)
	-	615	474 (77.1)	34 (5.5)	107 (17.4)
Placebo	+	226	180 (79.6)	18 (8.0)	28 (12.4)
	-	505	381 (75.4)	45 (8.9)	79 (15.6)
Total	+	339	275 (81.1)	23 (6.8)	41 (12.1)
	-	1120	855 (76.3)	79 (7.1)	186 (16.6)
	<b>All</b>	<b>1459</b>	<b>1130 (77.5)</b>	<b>102 (7.0)</b>	<b>227 (15.6)</b>

All: patients with either positive or negative local follow-up read included.

**TABLE 7 Overall distribution of outcomes for the central re-read of the follow-up scans, by randomised treatment for each trial**

Trial number and randomised treatment	Local follow-up result	Number of patients	Number (%) of patients with central follow-up result		
			Outcome of + or -	Indeterminate	Missing

<b>Trial IL0023</b>					
CASODEX	+	21	18 (85.7)	0 (0.0)	3 (14.3)
	-	221	160 (72.4)	11 (5.0)	50 (22.6)
Placebo	+	15	12 (80.0)	2 (13.3)	1 (6.7)
	-	179	143 (79.9)	11 (6.1)	25 (14.0)
<b>Trial IL0024</b>					
CASODEX	+	60	48 (80.0)	5 (8.3)	7 (11.7)
	-	213	165 (77.5)	15 (7.0)	33 (15.5)
Placebo	+	116	92 (79.3)	7 (6.0)	17 (14.7)
	-	187	132 (70.6)	18 (9.6)	37 (19.8)
<b>Trial IL0025</b>					
CASODEX	+	32	29 (90.6)	0 (0.0)	3 (9.4)
	-	181	149 (82.3)	8 (4.4)	24 (13.3)
Placebo	+	95	76 (80.0)	9 (9.5)	10 (10.5)
	-	139	106 (76.3)	16 (11.5)	17 (12.2)

**TABLE 8 Follow-up scan re-read results, by randomised treatment for each trial: estimate of proportion of positive and negative follow-up scans (assumption I)**

Trial number and randomised treatment	Local follow-up result	Number of patients	Number (%) of patients with central follow-up result	
			+	-
<b>Trial IL0023</b>				
CASODEX	+	18	13 (72.2)	5 (27.8)
	-	160	10 (6.3)	150 (93.7)
Placebo	+	12	8 (66.7)	4 (33.3)
	-	143	7 (4.9)	136 (95.1)
<b>Trial IL0024</b>				
CASODEX	+	48	33 (68.8)	15 (31.2)
	-	165	12 (7.3)	153 (92.7)
Placebo	+	92	70 (76.1)	22 (23.9)
	-	132	10 (7.6)	122 (92.4)
<b>Trial IL0025</b>				
CASODEX	+	29	27 (93.1)	2 (6.9)
	-	149	9 (6.0)	140 (94.0)

Placebo	+	76	66 (86.8)	10 (13.2)
	-	106	5 (4.7)	101 (95.3)

Assumption I: data presented are from patients with a baseline scan determined to be positive or negative,

ie, excluding outcomes of indeterminate or missing; equivalent to having included indeterminate and missing as being proportionately distributed.

**TABLE 9 Re-estimated hazard ratios (CASODEX vs. placebo treatment effect) for TTP based on the outcome of the central re-read**

Percentage of non-bone scan events retained in the calculation	'Adjusted' hazard ratio <sup>a</sup>		Original treatment effect for TTP analysed without event times
	Assumption I	Assumption II	
<b>100</b>	<b>0.723</b>	<b>0.745</b>	<b>0.63</b>
75	0.745	0.768	0.65
50	0.768	0.794	0.67
25	0.794	0.823	0.70
0	0.822	0.855	0.73

<sup>a</sup> A ratio <1 indicates a benefit for CASODEX compared with placebo.

Assumption I: patients with an outcome of indeterminate or missing were assumed to be distributed proportionally as positive and negative.

Assumption II: patients with an outcome of indeterminate were assumed to be negative, and those with outcome of missing were assumed to be distributed proportionally as positive and negative.

TABLE 10

RETROSPECTIVE CENTRAL RE-READ OF BONE SCANS FROM THE EPC PROGRAMME  
LOGISTIC ANALYSIS OF THE ESTIMATED NUMBER OF OBJECTIVE PROGRESSION EVENTS

TRIALS 0023, 0024 AND 0025 COMBINED

OTHER OBJ PROGS RETAINED AS AN EVENT = 0%

	ESTIMATE OF HAZARD RATIO #	

		(I)	(II)	
		+	+	
ANALYSIS OF:				
ESTIMATED NUMBER OF EVENTS		0.822	0.855	
		+	+	
LOWER 90% CL OF ESTIMATED NUMBER OF EVENTS		0.791	0.825	
		+	+	
UPPER 90% CL OF ESTIMATED NUMBER OF EVENTS		0.841	0.874	